

1 Amendments to the Claims

2
3 1. (Currently Amended) A method of identifying one or more
4 portions of a document, the method comprising:

5 identifying a plurality of visual blocks in the document based on, at least, a
6 document model of the document;

7 detecting, distinct from the plurality of visual blocks, one or more
8 separators of the document based on, at least, one or more characteristics of at
9 least one of the plurality of visual blocks between the visual blocks of the plurality
10 of visual blocks, wherein detecting the one or more separators comprises:

11 ~~initializing a separator list that includes one or more possible~~
12 ~~separators between the visual blocks;~~

13 ~~analyzing, for the visual blocks, whether the visual block overlaps a~~
14 ~~separator of the separator list, and if so how the visual block overlaps the~~
15 ~~separator; and~~

16 ~~determining how to treat the separator based on whether the visual~~
17 ~~block overlaps the separator, and if so how the visual block overlaps the separator;~~
18 ~~and~~

19 constructing, based at least in part on the plurality of visual blocks and the
20 one or more separators, a content structure for the document, wherein the content
21 structure identifies the different visual blocks as different portions of semantic
22 content of the document.

23
24 2. (Canceled)
25

1 3. (Previously Presented) A method as recited in claim 1, wherein
2 the document is described by a tree structure having a plurality of nodes, and
3 wherein identifying the plurality of visual blocks in the document comprises:

4 identifying a group of candidate nodes of the plurality of nodes;
5 for the respective nodes in the group of candidate nodes:
6 determining whether the node can be divided, and
7 if the node cannot be divided, then identifying the node as
8 representing a visual block.

9
10 4. (Original) A method as recited in claim 3, wherein if the node
11 cannot be divided, then setting a degree of coherence for the visual block
12 represented by the node.

13
14 5. (Original) A method as recited in claim 3, wherein if the node
15 cannot be divided, then removing the node from the group of candidate nodes.

16
17 6. (Canceled)

18
19 7. (Original) A method as recited in claim 3, wherein determining
20 whether the node can be divided comprises determining that the node can be
21 divided if a background color of the node is different from a background color of a
22 child of the node.

1 8. (Original) A method as recited in claim 3, further comprising
2 checking whether the node has a child having a width and height greater than zero,
3 and if the node has no child having a width and height greater than zero then
4 removing the node from the group of candidate nodes.

5
6 9. (Original) A method as recited in claim 3, wherein determining
7 whether the node can be divided comprises determining that the node can be
8 divided if a size of the node is at least a threshold amount greater than a sum of
9 sizes of children nodes of the node.

10
11 10. (Canceled)

12
13 11. (Original) A method as recited in claim 1, wherein the document
14 is described by a tree structure having a plurality of nodes, and wherein identifying
15 the plurality of visual blocks in the document comprises identifying different
16 visual blocks based at least in part on HyperText Markup Language (HTML) tags
17 of the plurality of nodes.

18
19 12. (Original) A method as recited in claim 1, wherein the document
20 is described by a tree structure having a plurality of nodes, and wherein identifying
21 the plurality of visual blocks in the document comprises identifying different
22 visual blocks based at least in part on background colors of the plurality of nodes.
23
24
25

1 13. (Original) A method as recited in claim 1, wherein the document
2 is described by a tree structure having a plurality of nodes, and wherein identifying
3 the plurality of visual blocks in the document comprises identifying different
4 visual blocks based at least in part on whether the plurality of nodes include text
5 and the sizes of the plurality of nodes.

6
7 14. (Currently Amended) A method as recited in claim 1, wherein:
8 the document has, at least, a horizontal direction and a vertical direction;
9 and

10 detecting the one or more separators comprises:

11 detecting one or more horizontal separators of the document ~~between~~
12 ~~the visual blocks~~; and

13 detecting one or more vertical separators of the document ~~between~~
14 ~~the visual blocks~~.

15
16 15. (Canceled)

17
18 16. (Currently Amended) A method as recited in claim 1, further
19 comprising determining to split a particular one of the separators into multiple
20 separators if one or more of the plurality of visual blocks is contained in the
21 particular separator.

17. (Currently Amended) A method as recited in claim 1, further comprising determining, if one or more of the plurality of visual blocks ~~eresses~~ overlap a particular one of the separators, to modify one or more parameters of the particular separator so that the one or more of the plurality of visual blocks no longer ~~eresses~~ overlap the particular separator.

18 – 19. (Canceled)

20. (Currently Amended) A method as recited in claim 1, further comprising determining to remove a particular one of the separators from the a separator list if one or more of the plurality of visual blocks ~~eevers~~ cover the particular separator.

21. (Original) A method as recited in claim 1, further comprising assigning, to each of the one or more separators, a weight based on characteristics of visual blocks on either side of the separator.

22. (Original) A method as recited in claim 21, wherein assigning the weight comprises assigning the weight based on a distance between two visual blocks on either side of the separator.

23. (Original) A method as recited in claim 21, wherein assigning the weight comprises assigning the weight based on whether the separator is at a same position as an <HR> HTML tag.

1 24. (Original) A method as recited in claim 21, wherein assigning the
2 weight comprises assigning the weight based on a font size used in two visual
3 blocks on either side of the separator.

4
5 25. (Original) A method as recited in claim 21, wherein assigning the
6 weight comprises assigning the weight based on a background color used in two
7 visual blocks on either side of the separator.

8
9 26. (Original) A method as recited in claim 1, further comprising:
10 checking whether each of the plurality of visual blocks satisfies a degree of
11 coherence threshold; and

12 for each of the plurality of visual blocks that does not satisfy the degree of
13 coherence threshold, identifying a new plurality of visual blocks in the visual
14 block, and repeating the detecting and constructing using the new plurality of
15 visual blocks.

16
17 27. (Original) A method as recited in claim 1, wherein constructing
18 the content structure comprises:

19 generating one or more virtual blocks based on the plurality of visual
20 blocks; and

21 including, in the content structure, the one or more virtual blocks.

22
23 28. (Original) A method as recited in claim 27, wherein generating
24 the one or more virtual blocks comprises generating the one or more virtual blocks
25 by combining two visual blocks of the plurality of visual blocks.

1
2 29. (Original) A method as recited in claim 27, further comprising:
3 determining a degree of coherence value for each of the one or more virtual
4 blocks.

5
6 30. (Original) A method as recited in claim 29, wherein determining
7 the degree of coherence value for a virtual block comprises determining the degree
8 of coherence value for the virtual block based at least in part on a weight of a
9 separator between two visual blocks used to generate the virtual block.

10
11 31. (Currently Amended) One or more computer readable media
12 having stored thereon a plurality of instructions that, when executed by one or
13 more processors of a device, causes the one or more processors to, at least:

14 identify visual blocks in a document based on, at least, a document model;
15 detect, distinct from the visual blocks, visual separators of the document
16 based on, at least, one or more characteristics of at least one of the visual blocks
17 ~~between the visual blocks, wherein instructions to detect visual separators~~
18 ~~comprise instructions to:~~

19 initialize a separator list that includes one or more possible visual
20 separators between the visual blocks;

21 analyze, for the visual blocks, whether the visual block overlaps a
22 separator of the separator list, and if so how the visual block overlaps the
23 separator, and
24
25

1 ~~determine how to treat the separator based on whether the visual~~
2 ~~block overlaps the separator, and if so how the visual block overlaps the separator;~~
3 and

4 construct, based at least in part on the visual blocks and the visual
5 separators, a content structure for the document that identifies regions of the
6 document that represent semantic content of the document.

7
8 32. (Original) One or more computer readable media as recited in
9 claim 31, wherein the document is described by a tree structure having a plurality
10 of nodes, and wherein the instructions that cause the one or more processors to
11 identify visual blocks in the document comprise instructions that cause the one or
12 more processors to:

13 identify a group of candidate nodes of the plurality of nodes;

14 for each node in the group of candidate nodes:

15 determine whether the node can be divided, and

16 if the node cannot be divided, then identify the node as representing
17 a visual block.

18
19 33. (Currently Amended) One or more computer readable media as
20 recited in claim 31, wherein:

21 the document has, at least, a horizontal direction and a vertical direction;

22 and

23 the instructions that cause the one or more processors to detect visual
24 separators comprise instructions that cause the one or more processors to, at least:
25

1 detect one or more horizontal separators of the document between
2 the visual blocks; and

3 detect one or more vertical separators of the document between the
4 visual blocks.

5
6 34. (Canceled)

7
8 35. (Original) One or more computer readable media as recited in
9 claim 31, wherein the instructions further cause the one or more processors to:

10 check whether each of the visual blocks satisfies a degree of coherence
11 threshold; and

12 for each of the visual blocks that does not satisfy the degree of coherence
13 threshold, identify new visual blocks in the visual block, and repeat the detection
14 and construction using the new visual blocks.

15
16 36 – 67. (Canceled)

17
18 68. (Currently Amended) A system comprising:

19 a visual block extractor, embodied at least in part in a computer readable
20 medium, to extract visual blocks from a document based on, at least, a document
21 model;

22 a visual separator detector, embodied at least in part in a computer readable
23 medium, coupled to receive the extracted visual blocks and configured to, at least,
24 detect, based on, at least, one or more characteristics of the extracted visual
25 blocks, one or more visual separators of the document between the extracted

1 visual blocks, wherein the visual separator detector detects the one or more visual
2 separators by:

3 initializing a separator list that includes one or more possible
4 separators between the visual blocks;

5 analyzing, for the visual blocks, whether the visual block overlaps a
6 separator of the separator list, and if so how the visual block overlaps the
7 separator; and

8 determining how to treat the separator based on whether the visual
9 block overlaps the separator, and if so how the visual block overlaps the separator;
10 and

11 a content structure constructor, embodied at least in part in a computer
12 readable medium, coupled to receive the extracted visual blocks and the detected
13 visual separators, and to use the extracted visual blocks and the detected visual
14 separators configured to, at least, construct a content structure for the document
15 based on, at least:

16 one or more of the extracted visual blocks; and

17 one or more of the visual separators.

18
19 69. (Original) A system as recited in claim 68, further comprising:

20 a document retrieval module to retrieve documents from a plurality of
21 documents based at least in part on the content structure constructed for one or
22 more of the plurality of documents.

1 70. (Original) A system as recited in claim 68, wherein the document
2 is described by a tree structure having a plurality of nodes, and wherein the visual
3 block extractor is to extract visual blocks from the document by:

4 identifying a group of candidate nodes of the plurality of nodes;

5 for each node in the group of candidate nodes:

6 determining whether the node can be divided, and

7 if the node cannot be divided, then identifying the node as
8 representing a visual block.

9
10 71. (Currently Amended) A system as recited in claim 68, wherein:

11 the document has, at least, a horizontal direction and a vertical direction;

12 and

13 the visual separator detector is further configured to, at least:

14 detect one or more horizontal separators of the document, between
15 ~~the visual blocks and;~~

16 detect one or more vertical separators of the document, between the
17 ~~visual blocks~~

18
19
20 72. (Canceled)

21
22 73. (Original) A system as recited in claim 68, wherein the content
23 structure constructor is further to:

24 check whether each of the plurality of visual blocks satisfies a degree of
25 coherence threshold; and

1 for each of the plurality of visual blocks that does not satisfy the degree of
2 coherence threshold, return the visual block to the visual block extractor to have a
3 new plurality of visual blocks extracted from the visual block, and further to have
4 the visual separator detector detect one or more visual separators using the new
5 plurality of visual blocks.

6
7 74. (Currently Amended) A system comprising:

8 means, embodied at least in part in a computer readable medium, for
9 identifying a plurality of visual blocks in ~~the a~~ document based on, at least, a
10 document model of the document;

11 means, embodied at least in part in a computer readable medium, for
12 detecting, distinct from the plurality of visual blocks, one or more separators of the
13 document based on, at least, one or more characteristics of at least one of the
14 plurality of visual blocks between the visual blocks of the plurality of visual
15 blocks, wherein the visual separator detector detects the one or more visual
16 separators by:

17 initializing a separator list that includes one or more possible
18 separators between the visual blocks,

19 analyzing, for the visual blocks, whether the visual block overlaps a
20 separator of the separator list, and if so how the visual block overlaps the
21 separator, and

22 determining how to treat the separator based on whether the visual
23 block overlaps the separator, and if so how the visual block overlaps the separator;
24 and
25

means, embodied at least in part in a computer readable medium, for constructing, based at least in part on the plurality of visual blocks and the one or more separators, a content structure for the document, wherein the content structure identifies the different visual blocks as different portions of semantic content of the document.

75. (Previously Presented) A system as recited in claim 74, wherein the document is described by a tree structure having a plurality of nodes, and wherein the means for identifying the plurality of visual blocks in the document comprises:

means, embodied at least in part in a computer readable medium, for identifying a group of candidate nodes of the plurality of nodes;

for each node in the group of candidate nodes:

means, embodied at least in part in a computer readable medium, for determining whether the node can be divided, and

means, embodied at least in part in a computer readable medium, for identifying, if the node cannot be divided, the node as representing a visual block.

76. (New) A method as recited in claim 1, wherein:
visual blocks are specified with respect to the document model; and
separators are specified with respect to the document as it would be displayed.

77. (New) A method as recited in claim 76, wherein the separator specification comprises a specification of a display area.

1
2 78. (New) A method as recited in claim 77, wherein the
3 specification of the display area comprises a specification of a start pixel and a
4 specification of an end pixel.
5

6 79. (New) A method as recited in claim 1, wherein detecting one
7 or more separators of the document comprises initializing a specification of an
8 initial separator to include a display area that would be occupied by the entire
9 document if it were displayed.
10

11 80. (New) A method as recited in claim 1, wherein detecting one
12 or more separators of the document comprises initializing a specification of an
13 initial separator to include a display area that would contain each of the plurality
14 of visual blocks if they were displayed.
15
16
17
18
19
20
21
22
23
24
25